

METHOD FOR QUEUED OVERLAP TRANSFER OF FILES

Insert

A1

FIELD OF THE INVENTION

The invention relates generally to the field of intelligent caching systems, and more particularly to a system and method for the transfer of files from a server machine to a client machine using an arbitrary transfer protocol, in which files are transferred in the background during the standard operation of a client-side software application.

BACKGROUND OF THE INVENTION

10 Many computer software programs provide functionality that allows a user to transfer files between a client and a server computer. These software programs typically provide such functionality in a straightforward fashion, allowing the user to select those files they want to transfer, and whether or not the files are to be sent to the server computer, or sent from the server computer to the client computer. One example of a typical software
15 program that provides such functionality is a File Transfer Protocol (FTP) program. The user can typically specify which FTP server they would like to connect to, and upon being connected, transfer files back and forth between the server and the client.

While some software programs provide file transfer functionality directly to the user,
20 there are others that perform file transfers in the background. These types of programs typically provide some other type of functionality, and the transfer of files is only a necessary part of providing the other functionality. An example of such a program is a so-called Web browser, examples of which include Microsoft Internet Explorer and

Netscape Navigator. The primary purpose of a Web browser is to allow the user to connect to other computers and view information provided in Hypertext Mark-up Language (HTML). Although the user has no explicit control over the transfer of files, the Web browser must request from the server those files it needs to display the information the user wants to see. All transfer of files takes place in the background. The user is typically oblivious to exactly what files are being transferred and in what order. Further, the user typically does not have control over this. Finally, typical web browsers load only those objects specified in the HTML document.

So called Web accelerators may define a further level of functionality, by preloading Web pages or other objects referenced on an active HTML document, and/or refreshing preferred Web pages previously visited and likely to be revisited. While some Web accelerators may provide a degree of user control, these rarely understand the content of the document and the inferred relationship between various documents within a set.

Whether files are transferred because of direct control of the user, or whether they are transferred in behalf of the user to provide additional functionality, there are several limitations to this approach. Firstly, the user typically does not have control over the order in which the files are transferred. Secondly, if the server computer must charge for the transfer of files, there is no way for the user to specify how many files are to be transferred. And thirdly, files are not automatically queued for transfer in an intelligent manner as to allow the user to make use of what he or she wants to in an organized fashion.

SUMMARY AND OBJECTS OF THE INVENTION

The present invention solves these issues by queuing up for the user files in an organized, pre-determined manner. The user is allowed to specify the number of files to be queued for transfer, and how many files the application software should "look ahead."

5

These functions are preferably integrated into a browser application, or as a plug-in or extension to the browser.

10

Because the process occurs predictably, the user may gain a sense of which actions have a high latency penalty and which are efficient.

15

Where an accounting system accounts for user "consumption" of documents, it is possible to contingently download documents in compressed format, and only account for their usage (or provide the appropriate decryption key) only when actually consumed (e.g., viewed, printed, analyzed, etc.). This system presumes, however, that the cost of document retrieval and transmission by the server is but a small operating cost, and that the user is willing to pay for this convenience based on documents or objects actually "consumed". This is particularly the case in the event of copyright content wherein an author or publisher must be compensated for the document.

20

On the other hand, where the document retrieval and transmission is substantial as compared to the overall value of the transaction, then it might be presumed that the cost is low, since database and telecommunication systems at present are quite efficient.

Thus, the user in that case might be very willing to pay for contingently downloaded content, even if not "consumed". This is particularly the case where the system employs the Internet to transmit relatively low-valued data records. By charging for such contingent downloads, the burden on the server may be equitably distributed, and abuses
5 of the finite resources are discouraged.

Further, by caching documents locally which are expected to be requested or required, the burden on the server may be more evenly distributed, permitting the average load on the server to be closer to the peak performance. Likewise, a user may predefine a
10 strategy, communicating with the server over a limited bandwidth link, with the negative impacts of server and communications latency minimized, assuming that the user must manually review and analyze the documents after downloading.

Typically, however, the caching strategy may be relatively simple: documents on a hit list
15 are downloaded sequentially, using the particular view format presently invoked to define the format of the records to download. If the format is changed while viewing a document, later documents in the list are cached in the new format, while prior documents are not, until the remainder of the list is exhausted. Thus, a linear review of the list is presumed. If this linear review is incorrect, the general strategy is for the user
20 to apply a different sort criterion or criteria to the list so that it better represents the desired order of viewing.

Upon connecting to a server computer, a list of files or documents is presented to the user. In an preferred embodiment, this list of files or documents would be presented in a list and the user would be able to select which file or document they want to view or want transferred. Typically, upon selecting the item for view or transfer, the present invention
5 requests from the server computer that it send the requested item.

Typically when a user is viewing such a list of items, he will want to view them sequentially. Thus it becomes important to decrease the time a user must wait when selecting the next item in the list for view or transfer.

10 The present invention provides a queuing mechanism whereby list items are queued for transfer sequentially ahead of the user.

Because responsiveness to the user is important, the present invention allows for the
15 cancellation of queued items, and allows for the continuation of queuing to take place in the case where the user skips several items in the list when making another request. In this process, the present invention cancels all requested items, requests the newly selected item, then continues by requesting additional items sequentially in the list after the newly selected item.

20 The present invention assumes that the user may be charged for items transferred by the server to the client. Thus, the present invention provides a setting that allows the user to specify whether or not all items in the list are transferred in advance of the user, or how

many items are to be queued in advance of where he/she is currently viewing or transferring. Typically if the user has specified that they want all of the files transferred, this can minimize connect time. If the user specifies some low value for this setting, then the connection time can be elongated, but charges for each item can be minimized.

5

The present invention assumes that the user may be charged for the time spent while connected to the server computer. Thus, the present invention provides support for settings that allow the user to specify the granularity or quality of the item to be sent in return. For example, if the items being transferred are images, the present invention provides support for a setting that dictates what resolution of image should be sent in return. The provision for these settings not only provides potential to decrease connect time, but also allow for the present invention to be more responsive in the cases where the user has selected less granularity.

10

15

The present invention allows the types of items being transferred to be arbitrary. That is to say, the items may be image files, text files, or any other type of file format. This file format is typically specific to the particular application. In the case of the preferred embodiment, the files transferred are modified Group 4 compressed TIFF images. Although Group 4 compressed TIFF images are used in the preferred embodiment, the choice of file format again is arbitrary and such a preferred embodiment could have been implemented using another image format, or a multiplicity of files providing the same information contained in the preferred embodiment's TIFF.

20

09691091 "101800

The present invention allows the user to specify the number of queue requests to be sent to the server at any one given time. This setting is different from the setting previously mentioned, responsible for allowing the user to specify the number of items queued in advance of the currently selected item. Typically on a high speed connection, the server is able to send to the client the queued files more quickly. Thus if the user decides to skip a number of items in the list when making a request, there are less queued files to be canceled. Because this setting allows the user to specify the number of queue requests to be sent to the server at any one given time, the user can adjust this setting to accommodate a slower or faster connection. If the user has a slower connection, then he/she will want to specify a smaller value for this setting so that less items have to be cancelled.

It is therefore an object of the invention to provide a system and method for prequeuing of files predicted to be desired by a user, through a telecommunications link, into a local cache, based on a list, wherein the prequeuing system is responsive to any change in the sequence of user review, such that predicted latencies for sequential file review from any given starting point are minimized.

The present invention also allows for an automated optimization strategy based on the throughput of the connection between the server and client, the apparent strategy for review of records by the user, the cost of the record downloads, the cost of on-line time, the value of the user's time, burden on the server, and/or possibly other factors. Using standard optimization techniques, including linear optimization, collaborative filtering,

fuzzy logic, neural networks and the like, it is possible to determine an optimal strategy that may be implemented. If conditions change, the strategy may be adaptive to these conditions and respond accordingly. Thus, while a relatively simple linear strategy of caching sequential records after the active record according to a list has been discussed in
5 detail herein, it should be understood that more complex schemes may be implemented using higher order optimization principles.

These and other objects will become apparent from a detailed review of the drawings and detailed description of the preferred embodiments.

10

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described with reference to the accompanying drawings, in which:

Fig. 1 shows a generic network, having a server and client connected through a
5 public network; and

Figs. 2A, 2B and 2C show a relation of the user review and prequeuing sequence according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The foregoing and other objects, features and advantages of the present invention will become more readily apparent to those skilled in the art to which the invention pertains upon reference to the following detailed description of one of the best modes for carrying out the invention, when considered in conjunction with the accompanying drawing in which preferred embodiments of the invention are shown and described by way of illustration, and not of limitation, wherein:

Fig. 1 shows a generic network. Typically, the server 1 has a high performance link 3 to the database storage subsystem 2. There may be latency in processing the request from the database storage subsystem 2, especially where the data is stored on optical disks or jukeboxes, magnetic tape, or other relatively slow storage systems. The telecommunications link 4 between the server 1 and the network 5 is also typically high performance, although there are instances where the bandwidth of this link is rate limiting.

The client system 7 has a telecommunications link 6 to the network, which is often rate limiting. Typically, the client side telecommunications link 6 is slower than the server side telecommunications link 4, but the server side telecommunications link 4 may be shared between multiple clients. In order to address the limitations of the client side telecommunications link 6, it is desired to prestore or queue files which may be

anticipated to be needed into the cache 8, which is typically a portion of a desktop computer hard drive. In comparison to the telecommunications links 4, 6, the cache 8 is rarely rate limiting.

- 5 A user interacts with the client 7 machine using a display 10, keyboard 11 and mouse 12.

When the user defines a search which returns a list 20, a sequence of documents are defined. The user will typically review these in sequence order, or may sort then according to a desire criterion. In either case, the user's review 23 will typically
10 commence with the first element on the list, denominated "1". The caching system will seek to download this file for the user 24. Assuming that the user's review takes longer than the download itself, the caching system will eventually get ahead 26 of the user 25, thus allowing the user to proceed without system latency delays.

- 15 If the user jumps ahead 27, 28, the caching system is reset from its prior status 29 to the new position of the user 30, canceling outstanding download requests.

In the field of patent searching, often a user must review each document on the list.

- Therefore, such shifts or jumps would be uncommon. Further, a professional reviewer
20 would likely procure a system having telecommunications bandwidth sufficient to avoid substantial latencies. Therefore, the preferred embodiment according to the present invention will effectively buffer the database storage subsystem 2, and

telecommunications link 4, 6 latencies to allow the user to achieve highest throughput after a short startup delay.

In some instances, the user does not sequentially review records on the list, for example
5 where a cursory examination is being made and decisions are made, for example, on the title of a record. In that case, the user may reset the options of the caching system to disable it, in order to reduce costs of downloading and to reduce the penalty incurred for interrupting ongoing downloads.

10 The foregoing description of the preferred embodiment of the invention has been presented for purposes of illustration and description and is not intended to be exhaustive or to limit the invention to the precise forms disclosed, since many modifications and variations are possible in light of the above teachings. Some modifications have been described in the specifications, and others may occur to those skilled in the art to which
15 the invention pertains. Therefore, the scope of the invention is to be defined solely by the claims that follow.